

RECEIVED

NOV 19 2003

TECH CENTER 1600/2900



1600

RAW SEQUENCE LISTING

DATE: 11/14/2003

PATENT APPLICATION: US/09/734,672A

TIME: 10:18:37

Input Set : N:\CrF4\11102003\I734672.raw
 Output Set: N:\CRF4\11142003\I734672A.raw

SEQUENCE LISTING

ENTERED

1 (1) GENERAL INFORMATION:
 2 (i) APPLICANT: Murphy, Patricia D.
 3 Allen, Antonette C.
 4 Alvares, Christopher P.
 5 Critz, Brenda S.
 6 Olson, Sheri J.
 7 Schelter, Denise B.
 8 Zeng, Bin
 9 (ii) TITLE OF INVENTION: Coding Sequences of the Human
 10 BRCA1 Gene
 11 (iii) NUMBER OF SEQUENCES: 72
 12 (iv) CORRESPONDENCE ADDRESS:
 13 (A) ADDRESSEE: Morgan Lewis & Bockius LLP
 14 (B) STREET: 1111 Pennsylvania Ave., N.W.
 15 (C) CITY: Washington
 16 (D) STATE: District of Columbia
 17 (E) COUNTRY: USA
 18 (F) ZIP: 20004
 19 (v) COMPUTER READABLE FORM:
 20 (A) MEDIUM TYPE: Floppy disk
 21 (B) COMPUTER: IBM PC compatible
 22 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 23 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
 24 (vi) CURRENT APPLICATION DATA:
 C--> 25 (A) APPLICATION NUMBER: US/09/734,672A
 C--> 26 (B) FILING DATE: 13-Dec-2000
 27 (vii) PRIOR APPLICATION DATA:
 W--> 28 (A) APPLICATION NUMBER: US 08/966,436
 29 (B) FILING DATE: 1997-11-07
 W--> 30 (A) APPLICATION NUMBER: US 08/598,591
 31 (B) FILING DATE: 1996-02-12
 32 (viii) ATTORNEY/AGENT INFORMATION:
 33 (A) NAME: Michael S. Tuscan
 34 (B) REGISTRATION NUMBER: 43,210
 35 (C) REFERENCE/DOCKET NUMBER: 44921-5055-02-US
 36 (ix) TELECOMMUNICATION INFORMATION:
 37 (A) TELEPHONE: 202-739-3000
 38 (B) TELEFAX: 202-739-3001
 39 (2) INFORMATION FOR SEQ ID NO: 1:
 40 (i) SEQUENCE CHARACTERISTICS:
 41 (A) LENGTH: 5711 base pairs
 42 (B) TYPE: nucleic acid

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/734,672A

DATE: 11/14/2003

TIME: 10:18:37

Input Set : N:\Crf4\11102003\I734672.raw

Output Set: N:\CRF4\11142003\I734672A.raw

```

43      (C) STRANDEDNESS: Not Relevant
44      (D) TOPOLOGY: linear
45      (ii) MOLECULE TYPE: cDNA
46      (vi) ORIGINAL SOURCE:
47          (A) ORGANISM: Homo sapiens
48          (B) STRAIN: BRCA1
49      (viii) POSITION IN GENOME:
50          (A) CHROMOSOME/SEGMENT: 17
51          (B) MAP POSITION: 17q21
52      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
53      AGCTCGCTGA GACTTCCTGG ACCCGCACC AGGCTGTGGG GTTTCTCAGA TAACTGGGCC      60
54      CCTGCGCTCA GGAGGCCTTC ACCCTCTGCT CTGGGTAAAG TTCATTGGAA CAGAAAGAAA      120
55      TGGATTTATC TGCTCTTCGC GTTGAAGAAG TACAAAATGT CATTAATGCT ATGCAGAAAA      180
56      TCTTAGAGTG TCCCATCTGT CTGGAGTTGA TCAAGGAACC TGTCTCCACA AAGTGTGACC      240
57      ACATATTTTG CAAATTTTGC ATGCTGAAAC TTCTCAACCA GAAGAAAGGG CCTTCACAGT      300
58      GTCCTTTATG TAAGAATGAT ATAACCAAAA GGAGCCTACA AGAAAGTACG AGATTTAGTC      360
59      AACTTGTTGA AGAGCTATTG AAAATCATTT GTGCTTTTCA GCTTGACACA GGTTTGGAGT      420
60      ATGCAAAACAG CTATAATTTT GCAAAAAAGG AAAATAACTC TCCTGAACAT CTAAAAGATG      480
61      AAGTTTCTAT CATCCAAAGT ATGGGCTACA GAAACCGTGC CAAAAGACTT CTACAGAGTG      540
62      AACCCGAAAA TCCTTCCTTG CAGGAAACCA GTCTCAGTGT CCAACTCTCT AACCTTGGAA      600
63      CTGTGAGAAC TCTGAGGACA AAGCAGCGGA TACAACCTCA AAAGACGTCT GTCTACATTG      660
64      AATTGGGATC TGATTCTTCT GAAGATACCG TTAATAAGGC AACTTATTGC AGTGTGGGAG      720
65      ATCAAGAATT GTTCAAAATC ACCCCTCAAG GAACCAGGGA TGAAATCAGT TTGGATTCTG      780
66      CAAAAAAGGC TGCTTGTGAA TTTTCTGAGA CGGATGTAAC AAATACTGAA CATCATCAAC      840
67      CCAGTAATAA TGATTTGAAC AACACTGAGA AGCGTGCAGC TGAGAGGCAT CCAGAAAAGT      900
68      ATCAGGGTAG TTCTGTTTCA AACTTGCATG TGGAGCCATG TGGCACAAAT ACTCATGCCA      960
69      GCTCATTACA GCATGAGAAC AGCAGTTTAT TACTCACTAA AGACAGAATG AATGTAGAAA      1020
70      AGGCTGAATT CTGTAATAAA AGCAAACAGC CTGGCTTAGC AAGGAGCCAA CATAACAGAT      1080
71      GGGCTGGAAG TAAGGAAACA TGTAATGATA GCGGACTCC CAGCACAGAA AAAAAGGTAG      1140
72      ATCTGAATGC TGATCCCCTG TGTGAGAGAA AAGAATGGAA TAAGCAGAAA CTGCCATGCT      1200
73      CAGAGAATCC TAGAGATACT GAAGATGTTT CTTGGATAAC ACTAAATAGC AGCATTCAGA      1260
74      AAGTTAATGA GTGGTTTTCC AGAAGTGATG AACTGTTAGG TTCTGATGAC TCACATGATG      1320
75      GGGAGTCTGA ATCAAATGCC AAAGTAGCTG ATGTATTGGA CGTTCTAAAT GAGGTAGATG      1380
76      AATATTCTGG TTCTTCAGAG AAAATAGACT TACTGGCCAG TGATCCTCAT GAGGCTTTAA      1440
77      TATGTAAAAG TGAAAGAGTT CACTCCAAAT CAGTAGAGAG TAATATTGAA GACAAAATAT      1500
78      TTGGGAAAAC CTATCGGAAG AAGGCAAGCC TCCCCAACTT AAGCCATGTA ACTGAAAATC      1560
79      TAATTATAGG AGCATTGTGTT ACTGAGCCAC AGATAATACA AGAGCGTCCC CTCACAAATA      1620
80      AATTAAAGCG TAAAAGGAGA CCTACATCAG GCCTTCATCC TGAGGATTTT ATCAAGAAAG      1680
81      CAGATTTGGC AGTTCAAAAG ACTCCTGAAA TGATAAATCA GGGAACCTAAC CAAACGGAGC      1740
82      AGAATGGTCA AGTGATGAAT ATTACTAATA GTGGTCATGA GAATAAAACA AAAGGTGATT      1800
83      CTATTCAGAA TGAGAAAAAT CCTAACCCAA TAGAATCACT CGAAAAAGAA TCTGCTTTCA      1860
84      AAACGAAAGC TGAACCTATA AGCAGCAGTA TAAGCAATAT GGAACCTCGAA TTAAATATCC      1920
85      ACAATTCAAA AGCACCTAAA AAGAATAGGC TGAGGAGGAA GTCTTCTACC AGGCATATTC      1980
86      ATGCGCTTGA ACTAGTAGTC AGTAGAAATC TAAGCCCACC TAATTGTACT GAATTGCAAA      2040
87      TTGATAGTTG TTCTAGCAGT GAAGAGATAA AGAAAAAAA GTACAACCAA ATGCCAGTCA      2100
88      GGCACAGCAG AAACCTACAA CTCATGGAAG GTAAAGAACC TGCAACTGGA GCCAAGAAGA      2160
89      GTAACAAGCC AAATGAACAG ACAAGTAAAA GACATGACAG TGATACTTTC CCAGAGCTGA      2220
90      AGTTAACAAA TGCACCTGGT TCTTTTACTA AGTGTTCAAA TACCAGTGAA CTTAAAGAAT      2280
91      TTGTCAATCC TAGCCTTCCA AGAGAAGAAA AAGAAGAGAA ACTAGAAACA GTTAAAGTGT      2340

```

RAW SEQUENCE LISTING

DATE: 11/14/2003

PATENT APPLICATION: US/09/734,672A

TIME: 10:18:37

Input Set : N:\Crf4\11102003\I734672.raw

Output Set: N:\CRF4\11142003\I734672A.raw

92	CTAATAATGC TGAAGACCCC AAAGATCTCA TGTTAAGTGG AGAAAGGGTT TTGCAAACCTG	2400
93	AAAGATCTGT AGAGAGTAGC AGTATTTTCAC TGGTACCTGG TACTGATTAT GGCACCTCAGG	2460
94	AAAGTATCTC GTTACTGGAA GTTAGCACTC TAGGGAAGGC AAAACAGAA CCAAATAAAT	2520
95	GTGTGAGTCA GTGTGCAGCA TTTGAAAACC CCAAGGGACT AATTCATGGT TGTTCCAAAG	2580
96	ATAATAGAAA TGACACAGAA GGCTTTAAGT ATCCATTGGG ACATGAAGTT AACCACAGTC	2640
97	GGGAAACAAG CATAGAAATG GAAGAAAGTG AACTTGATGC TCAGTATTTG CAGAATACAT	2700
98	TCAAGGTTTC AAAGCGCCAG TCATTTGCTC TGTTTTCAAA TCCAGGAAAT GCAGAAGAGG	2760
99	AATGTGCAAC ATTCTCTGCC CACTCTGGGT CCTTAAAGAA ACAAAGTCCA AAAGTCACTT	2820
100	TTGAATGTGA ACAAAGGAA GAAATCAAG GAAAGAATGA GTCTAATATC AAGCCTGTAC	2880
101	AGACAGTTAA TATCACTGCA GGCTTTCCTG TGTTTGGTCA GAAAGATAAG CCAGTTGATA	2940
102	ATGCCAAATG TAGTATCAAA GGAGGCTCTA GGTTTGTCT ATCATCTCAG TTCAGTGGCA	3000
103	ACGAAACTGG ACTCATTACT CCAAATAAAC ATGGACTTTT ACAAACCCA TATCGTATAC	3060
104	CACCACTTTT TCCCATCAAG TCATTTGTGA AAATAAATG TAAGAAAAAT CTGCTAGAGG	3120
105	AAAACCTTGA GGAACATTCA ATGTCACCTG AAAGAGAAAT GGGAAATGAG AACATTCCAA	3180
106	GTACAGTGAG CACAATTAGC CGTAATAACA TTAGAGAAAA TGTTTTTAAA GGAGCCAGCT	3240
107	CAAGCAATAT TAATGAAGTA GGTTCAGTA CTAATGAAGT GGGCTCCAGT ATTAATGAAA	3300
108	TAGGTTCCAG TGATGAAAAC ATTCAAGCAG AACTAGGTAG AAACAGAGGG CCAAATTGA	3360
109	ATGCTATGCT TAGATTAGGG GTTTTGCAAC CTGAGGTCTA TAAACAAAGT CTTCTGGAA	3420
110	GTAATTGTAA GCATCCTGAA ATAAAAAGC AAGAATATGA AGAAGTAGTT CAGACTGTGA	3480
111	ATACAGATTT CTCTCCATAT CTGATTTTCTG ATAACCTAGA ACAGCCTATG GGAAGTAGTC	3540
112	ATGCATCTCA GGTTTGTTCT GAGACACCTG ATGACCTGTT AGATGATGGT GAAATAAAGG	3600
113	AAGATACTAG TTTTGCTGAA AATGACATTA AGGAAAGTTC TGCTGTTTTT AGCAAAAGCG	3660
114	TCCAGAGAGG AGAGCTTAGC AGGAGTCCTA GCCCTTTCAC CCATACACAT TTGGCTCAGG	3720
115	GTTACCGAAG AGGGGCCAAG AAATTAGAGT CCTCAGAAGA GAACTTATCT AGTGAGGATG	3780
116	AAGAGCTTCC CTGCTTCCAA CACTTGTTAT TTGGTAAAGT AAACAATATA CTTTCTCAGT	3840
117	CTACTAGGCA TAGCACCGTT GCTACCGAGT GTCTGTCTAA GAACACAGAG GAACATTTAT	3900
118	TATCATTGAA GAATAGCTTA AATGACTGCA GTAACCAGGT AATATTGGCA AAGGCATCTC	3960
119	AGGAACATCA CCTTAGTGAG GAAACAAAAT GTTCTGCTAG CTTGTTTTCT TCACAGTGCA	4020
120	GTGAATTGGA AGACTTGACT GCAAATACAA ACACCCAGGA TCCTTTCTTG ATTGGTTCTT	4080
121	CCAAACAAAT GAGGCATCAG TCTGAAAGCC AGGGAGTTGG TCTGAGTGAC AAGGAATTGG	4140
122	TTTCAGATGA TGAAGAAAGA GGAACGGGCT TGGAAGAAAA TAATCAAGAA GAGCAAAGCA	4200
123	TGGATTCAAA CTTAGGTGAA GCAGCATCTG GGTGTGAGAG TGAAACAAGC GTCTCTGAAG	4260
124	ACTGCTCAGG GCTATCCTCT CAGAGTGACA TTTTAACCAC TCAGCAGAGG GATACCATGC	4320
125	AACATAACCT GATAAAGCTC CAGCAGGAAA TGGCTGAACT AGAAGCTGTG TTAGAACAGC	4380
126	ATGGGAGCCA GCCTTCTAAC AGCTACCCTT CCATCATAAG TGAATCCTCT GCCCTTGAGG	4440
127	ACCTGCGAAA TCCAGAACAA AGCACATCAG AAAAAGCAGT ATTAACCTCA CAGAAAAGTA	4500
128	GTGAATACCC TATAAGCCAG AATCCAGAAG GCCTTTCTGC TGACAAGTTT GAGGTGTCTG	4560
129	CAGATAGTTC TACCAGTAAA AATAAAGAAC CAGGAGTGGA AAGGTCATCC CTTTCTAAAT	4620
130	GCCCATCATT AGATGATAGG TGGTACATGC ACAGTTGCTC TGGGAGTCTT CAGAATAGAA	4680
131	ACTACCCATC TCAAGAGGAG CTCATTAAGG TTGTTGATGT GGAGGAGCAA CAGCTGGAAG	4740
132	AGTCTGGGCC ACACGATTTG ACGGAAACAT CTTACTTGCC AAGGCAAGAT CTAGAGGGAA	4800
133	CCCCTTACCT GGAATCTGGA ATCAGCCTCT TCTCTGATGA CCCTGAATCT GATCCTTCTG	4860
134	AAGACAGAGC CCCAGAGTCA GCTCGTGTG GCAACATACC ATCTTCAACC TCTGCATTGA	4920
135	AAGTTCCCCA ATTGAAAGTT GCAGAATCTG CCCAGGGTCC AGCTGCTGCT CATACTACTG	4980
136	ATACTGCTGG GTATAATGCA ATGGAAGAAA GTGTGAGCAG GGAGAAGCCA GAATTGACAG	5040
137	CTTCAACAGA AAGGGTCAAC AAAAGAATGT CCATGGTGGT GTCTGGCCTG ACCCCAGAAG	5100
138	AATTTATGCT CGTGTACAAG TTTGCCAGAA AACACCACAT CACTTTAACT AATCTAATTA	5160
139	CTGAAGAGAC TACTCATGTT GTTATGAAAA CAGATGCTGA GTTTGTGTGT GAACGGACAC	5220
140	TGAAATATTT TCTAGGAATT GCGGGAGGAA AATGGGTAGT TAGCTATTC TGGGTGACCC	5280

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/734,672A

DATE: 11/14/2003

TIME: 10:18:37

Input Set : N:\Crf4\11102003\I734672.raw

Output Set: N:\CRF4\11142003\I734672A.raw

```

141 AGTCTATTAA AGAAAGAAAA ATGCTGAATG AGCATGATTT TGAAGTCAGA GGAGATGTGG 5340
142 TCAATGGAAG AAACCACCAA GGTCCAAAGC GAGCAAGAGA ATCCCAGGAC AGAAAGATCT 5400
143 TCAGGGGGCT AGAAATCTGT TGCTATGGGC CCTTCACCAA CATGCCCACA GATCAACTGG 5460
144 AATGGATGGT ACAGCTGTGT GGTGCTTCTG TGGTGAAGGA GCTTTCATCA TTCACCCTTG 5520
145 GCACAGGTGT CCACCCAATT GTGGTTGTGC AGCCAGATGC CTGGACAGAG GACAATGGCT 5580
146 TCCATGCAAT TGGGCAGATG TGTGAGGCAC CTGTGGTGAC CCGAGAGTGG GTGTTGGACA 5640
147 GTGTAGCACT CTACCACTGC CAGGAGCTGG ACACCTACCT GATACCCCAG ATCCCCCACA 5700
148 GCCACTACTG A 5711
150 (2) INFORMATION FOR SEQ ID NO: 2:
151 (i) SEQUENCE CHARACTERISTICS:
152 (A) LENGTH: 1863 amino acids
153 (B) TYPE: amino acid
154 (C) STRANDEDNESS: Not Relevant
155 (D) TOPOLOGY: Unknown
156 (ii) MOLECULE TYPE: protein
157 (vi) ORIGINAL SOURCE:
158 (A) ORGANISM: Homo sapiens
159 (B) STRAIN: BRCA1
160 (viii) POSITION IN GENOME:
161 (A) CHROMOSOME/SEGMENT: 17
162 (B) MAP POSITION: 17q21
163 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
164 Met Asp Leu Ser Ala Leu Arg Val Glu Val Gln Asn Val Ile Asn
165 1 5 10 15
166 Ala Met Gln Lys Ile Leu Glu Cys Pro Ile Cys Leu Glu Leu Ile Lys
167 20 25 30
168 Glu Pro Val Ser Thr Lys Cys Asp His Ile Phe Cys Lys Phe Cys Met
169 35 40 45
170 Leu Lys Leu Leu Asn Gln Lys Lys Gly Pro Ser Gln Cys Pro Leu Cys
171 50 55 60
172 Lys Asn Asp Ile Thr Lys Arg Ser Leu Gln Glu Ser Thr Arg Phe Ser
173 65 70 75 80
174 Gln Leu Val Glu Glu Leu Leu Lys Ile Ile Cys Ala Phe Gln Leu Asp
175 85 90 95
176 Thr Gly Leu Glu Tyr Ala Asn Ser Tyr Asn Phe Ala Lys Lys Glu Asn
177 100 105 110
178 Asn Ser Pro Glu His Leu Lys Asp Glu Val Ser Ile Ile Gln Ser Met
179 115 120 125
180 Gly Tyr Arg Asn Arg Ala Lys Arg Leu Leu Gln Ser Glu Pro Glu Asn
181 130 135 140
182 Pro Ser Leu Gln Glu Thr Ser Leu Ser Val Gln Leu Ser Asn Leu Gly
183 145 150 155 160
184 Thr Val Arg Thr Leu Arg Thr Lys Gln Arg Ile Gln Pro Gln Lys Thr
185 165 170 175
186 Ser Val Tyr Ile Glu Leu Gly Ser Asp Ser Ser Glu Asp Thr Val Asn
187 180 185 190
188 Lys Ala Thr Tyr Cys Ser Val Gly Asp Gln Glu Leu Leu Gln Ile Thr
189 195 200 205
190 Pro Gln Gly Thr Arg Asp Glu Ile Ser Leu Asp Ser Ala Lys Lys Ala

```


VERIFICATION SUMMARY

PATENT APPLICATION: US/09/734,672A

DATE: 11/14/2003

TIME: 10:18:38

Input Set : N:\Crf4\11102003\I734672.raw

Output Set: N:\CRF4\11142003\I734672A.raw

L:25 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:26 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:30 M:238 W: Alpha Fields not Ordered, Reordered [(A) APPLICATION NUMBER:] of (1)(vii)